

## Supplementary material for the Epidemiological and phylogenetic analyses of public SARS-CoV-2 data from Malawi

SARS-Cov-2 lineages grouped into variants and named according to names given by the World Health Organisation (WHO).

COVID-19 variant				
Alpha (6 sequences)	Beta (492 sequences)	Delta (569 sequences)	Omicron (273 sequences)	Other (82 sequences)
B.1.1.7	B.1.351 B.1.36	AY.104 AY.116 AY.122 AY.30 AY.37 AY.38 AY.45 AY.46 AY.6 B.1.617.2	BA.1 BA.1.1 BA.1.13 BA.1.14 BA.1.14.2 BA.1.17 BA.1.17.2 BA.1.18 BA.1.19 BA.1.21 BA.1.9 BA.2 BA.2.65 BA.4 BA.4.1 BA.4.1.9 BA.4.6.1 BA.4.7 BA.5 BA.5.2 BQ.1.1	A.23.1 B B.1 B.1.1 B.1.1.1 B.1.1.33 B.1.1.375 B.1.1.412 B.1.1.448 B.1.1.54 B.1.177 B.6

The collected sequences in fasta format were aligned to identify regions of similarities. The sequence alignment process used MAFFT, a multiple alignment program for amino acid or nucleotide sequences, which we coded in python-jupyter notebook. The aligned sequences together with the reference sequence of Wuhan city, China, were later processed into evolution trees using IQ-TREE, a stochastic algorithm that builds phylogenetic trees by the maximum likelihood method [1]. IQ-TREE, which is an efficient, open-source software for phylogenomic inference was chosen for its efficiency in selecting the best model for tree reconstruction. The tree was rooted with a reference sequence of Wuhan using a model  $GTR + F + I + I + R5$  that emerged as the best model through the model finder [2]. TreeTime version 0.82 [3] was used to visualise the reconstructed national phylogenetic tree on a time scale while the trees

for individual variants were visualised and coloured by region using an online tool for displaying, annotating and managing trees called an Interactive Tree Of Life (iTOL).

## References

1. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*. 2020;37(5):1530–1534.
2. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*. 2017;14(6):587–589.
3. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus evolution*. 2018;4(1):vex042.